Pergamon

Computers Math. Applic. Vol. 31, No. 8, pp. 31–36, 1996

0898-1221/96 \$15.00 + 0.00

S0898-1221(96)00028-4

Further Analysis of Stability for Lambert's Method Based on Euler's Rule

H. HAYASHI
Department of Computer Science, University of Toronto
Toronto, Ontario, Canada M5S 1A4

T. MITSUI
Graduate School of Human Informatics, Nagoya University
Nagoya 464-01, Japan

(Received and accepted September 1995)

Abstract—Lambert [1] proposed a one-step method based on Euler's rule which effectively copes with ordinary differential equations. In the paper, he proved its stability for the equation of the type $\boldsymbol{y}' = \boldsymbol{A}\boldsymbol{y}$, where $\boldsymbol{A}$ is a real symmetric matrix. We extend the concept and prove the stability when $\boldsymbol{A}$ is a real normal matrix.

Keywords—Stability analysis, Lambert's method, Nonlinear method, Numerical ODEs, Real normal matrix.

1. INTRODUCTION

We consider the initial value problem

$$\boldsymbol{y}' = \boldsymbol{f}(x, \boldsymbol{y}), \qquad \boldsymbol{y}(a) = \boldsymbol{y}_0, \tag{1}$$

where $\boldsymbol{y}$, $\boldsymbol{f} \in \boldsymbol{R}^m$ and $x \in [a, b]$. Lambert [1] has proposed a one-step method with self-adjusted steplength which is based on Euler's rule. It is given by

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h_n \boldsymbol{f}_n, \qquad \text{with } h_n = \min\left(h_0, h_n^*\right), \quad h_n^* = \frac{\boldsymbol{f}_n^\top \boldsymbol{f}_n}{|\boldsymbol{f}_n^\top \boldsymbol{A} \boldsymbol{f}_n|}, \tag{2}$$

where h_0 is the initial steplength and $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}|_n$ is used instead of $\boldsymbol{A}$ when equation (1) is nonlinear. The advantage of this method is that we can obtain a cheap low-accuracy solution to a stiff system of ordinary differential equations.

In fact, in his paper [1], Lambert proved the following theorem.

THEOREM 1. *Let the method (2) be applied to the linear problem*

$$\boldsymbol{y}' = \boldsymbol{A}\boldsymbol{y}, \qquad \boldsymbol{y}(x_0) = \boldsymbol{y}_0,$$

where $\boldsymbol{A}$ is a real symmetric $m \times m$ matrix with eigenvalues λ_t, $t = 1, 2, \ldots, m$, satisfying

$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_m < 0.$$

The authors thank the referee whose comments enable us to improve this article.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Then as $n \to \infty$, $\boldsymbol{y}_n \to 0$ monotonically in the L_2-norm.

A lemma below, which was also proved in [1], supports the proof of Theorem 1.

LEMMA 1. *Let $\boldsymbol{B}$ be a real symmetric $m \times m$ matrix with eigenvalues α_t, $t = 1, 2, \dots, m$, satisfying*

$$\alpha_1 \le \alpha_2 \le \cdots \le \alpha_m < 0.$$

Then for any $\boldsymbol{w} \in \boldsymbol{R}^m$, $\boldsymbol{w} \ne \boldsymbol{0}$, there hold

$$\alpha_1 \le \frac{\boldsymbol{w}^\top \boldsymbol{B} \boldsymbol{w}}{\boldsymbol{w}^\top \boldsymbol{w}} \le \alpha_m, \quad \frac{\left(\boldsymbol{w}^\top \boldsymbol{w}\right)^2}{\left(\boldsymbol{w}^\top \boldsymbol{B} \boldsymbol{w}\right)\left(\boldsymbol{w}^\top \boldsymbol{B}^{-1} \boldsymbol{w}\right)} \le 1.$$

Thus the statement of Theorem 1 is only valid for the linear system with real symmetric matrix coefficients. This means Lambert's method can give a stable and cheap approximation for the exponentially decreasing solutions without oscillation. But our preliminary numerical observations suggest a possibility to extend the stability property of the method over the real symmetric case.

Our aim here is to extend Theorem 1 when $\boldsymbol{A}$ is a real normal matrix. In Section 4, we will show that some numerical experiments confirm our theorem.

2. A KEY LEMMA

Throughout Sections 2 and 3, we use the following assumption on matrices.

ASSUMPTION A. *A real normal $m \times m$ matrix has its eigenvalues λ_i for $i = 1, 2, \dots, m$ such that*

$$\begin{aligned} &\lambda_{2i-1} = \alpha_i + \sqrt{-1} \cdot \beta_i, \quad \lambda_{2i} = \bar{\lambda}_{2i-1}, && \text{for } 1 \le i \le k, \\ &\lambda_i \in \boldsymbol{R}, && \text{for } 2k+1 \le i \le m, \end{aligned}$$

and $\Re\lambda_i < 0$ for $1 \le i \le m$.

LEMMA 2. *Assume that a matrix $\boldsymbol{B}$ satisfies Assumption A. Let $l = \min \Re\lambda_i$ and $L = \max \Re\lambda_i$; then for any $\boldsymbol{w} = [w_1, w_2, \dots, w_m]^\top \in \boldsymbol{R}^m$ and $\boldsymbol{w} \ne \boldsymbol{0}$, the following inequality holds:*

$$l \le \frac{\boldsymbol{w}^\top \boldsymbol{B} \boldsymbol{w}}{\boldsymbol{w}^\top \boldsymbol{w}} = \frac{(1/2)\boldsymbol{w}^\top \left(\boldsymbol{B}^\top + \boldsymbol{B}\right) \boldsymbol{w}}{\boldsymbol{w}^\top \boldsymbol{w}} \le L. \tag{3}$$

Furthermore, the inequality

$$\frac{\left(\boldsymbol{w}^\top \boldsymbol{w}\right)^2}{\left(\boldsymbol{w}^\top \boldsymbol{B} \boldsymbol{w}\right) \cdot (1/2)\boldsymbol{w}^\top \left(\boldsymbol{B}^{-1} + \boldsymbol{B}^{-\top}\right) \boldsymbol{w}} \le 2 - \delta \tag{4}$$

holds with a certain positive δ for any $\boldsymbol{w} \in \boldsymbol{R}^m$ if and only if the inequality $|\beta_i / \alpha_i| < 1$ holds for all i $(1 \le i \le k)$.

PROOF. Since the matrix $\boldsymbol{B}$ is real and normal, there exists [2, p. 105] a real orthogonal matrix $\boldsymbol{Q}$ such that

$$\boldsymbol{Q}^\top \boldsymbol{B} \boldsymbol{Q} = \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_1 & & & & \boldsymbol{0} \\ & \boldsymbol{\Lambda}_2 & & & \\ & & \ddots & & \\ & & & \boldsymbol{\Lambda}_k & \\ \boldsymbol{0} & & & & \tilde{\boldsymbol{\Lambda}} \end{bmatrix},$$

where the submatrices are given by

$$\boldsymbol{\Lambda}_i = \begin{bmatrix} \alpha_i & \beta_i \\ -\beta_i & \alpha_i \end{bmatrix}, \qquad \tilde{\boldsymbol{\Lambda}} = \begin{bmatrix} \lambda_{2k+1} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \lambda_m \end{bmatrix}.$$

Let

$$w = Qu, \qquad u = [u_1, u_2, \dots, u_m]^\top$$

and $v_i = u_{2i-1}^2 + u_{2i}^2$ for $1 \le i \le k$; then we have the following identities:

$$w^\top B w = u^\top \Lambda u = \sum_{i=1}^{k} \alpha_i v_i + \sum_{i=2k+1}^{m} \lambda_i u_i^2, \tag{5}$$

$$\begin{aligned}\frac{1}{2} w^\top \left(B^\top + B\right) w &= \frac{1}{2} u^\top \left(Q^\top B^\top Q + Q^\top B Q\right) u \\ &= \frac{1}{2} u^\top \left(\Lambda^\top + \Lambda\right) u \\ &= \sum_{i=1}^{k} \alpha_i v_i + \sum_{i=2k+1}^{m} \lambda_i u_i^2 = w^\top B w, \end{aligned} \tag{6}$$

$$\begin{aligned}\frac{1}{2} w^\top \left(B^{-1} + B^{-\top}\right) w &= \frac{1}{2} u^\top \left(Q^\top B^{-1} Q + Q^\top B^{-\top} Q\right) u \\ &= \frac{1}{2} u^\top \left(\Lambda^{-1} + \Lambda^{-\top}\right) u \\ &= \frac{1}{2} \left[\sum_{i=1}^{k} \frac{2\,(1/\alpha_i)}{1 + (\beta_i/\alpha_i)^2} v_i + \sum_{i=2k+1}^{m} \frac{2}{\lambda_i} u_i^2\right], \end{aligned} \tag{7}$$

$$w^\top w = u^\top u = \sum_{i=1}^{m} u_i^2. \tag{8}$$

Let $\sigma_{2i-1} = \sigma_{2i} = \alpha_i$ for $1 \le i \le k$ and $\sigma_i = \lambda_i$ for $2k+1 \le i \le m$. Equations (5) and (8) yield

$$\frac{w^\top B w}{w^\top w} = \frac{\sum_{i=1}^{k} \alpha_i v_i + \sum_{i=2k+1}^{m} \lambda_i u_i^2}{\sum_{i=1}^{m} u_i^2} = \frac{\sum_{i=1}^{m} \sigma_i u_i^2}{\sum_{i=1}^{m} u_i^2},$$

and the desired result (3) follows from (6) and the fact that

$$l \sum_{i=1}^{m} u_i^2 \le \sum_{i=1}^{m} \sigma_i u_i^2 \le L \sum_{i=1}^{m} u_i^2.$$

Using (5), (7), and (8), we obtain

$$\begin{aligned} &2\left(w^\top B w\right) \frac{1}{2} w^\top \left(B^{-1} + B^{-\top}\right) w - \left(w^\top w\right)^2 \\ &= \left[\sum_{i=1}^{k} \alpha_i v_i + \sum_{i=2k+1}^{m} \lambda_i u_i^2\right] \left[\sum_{i=1}^{k} \frac{2\,(1/\alpha_i)}{1 + (\beta_i/\alpha_i)^2} v_i + \sum_{i=2k+1}^{m} \frac{2}{\lambda_i} u_i^2\right] - \left(\sum_{i=1}^{m} u_i^2\right)^2. \end{aligned}$$

From the assumption, we can take a positive number ε (< 1) such that the inequalities $|\beta_i/\alpha_i| \le 1 - \varepsilon$ hold for $1 \le i \le k$. The substitution of σ_i into the above identity imply

$$\begin{aligned} &2\left(w^\top B w\right) \frac{1}{2} w^\top \left(B^{-1} + B^{-\top}\right) w - \left(w^\top w\right)^2 \\ &\ge \left[\sum_{i=1}^{k} \alpha_i v_i + \sum_{i=2k+1}^{m} \lambda_i u_i^2\right] \left[\sum_{i=1}^{k} \frac{2}{1 + (1-\varepsilon)^2} \left(\frac{1}{\alpha_i}\right) v_i + \sum_{i=2k+1}^{m} \frac{2}{\lambda_i} u_i^2\right] - \left(\sum_{i=1}^{m} u_i^2\right)^2 \\ &= \left(\sum_{i=1}^{m} \sigma_i u_i^2\right) \left(\sum_{i}^{m} \frac{1}{\sigma_i} u_i^2\right) - \left(\sum_{i=1}^{m} u_i^2\right)^2 + R_2 \\ &= \sum_{i}^{m} u_i^4 + \sum_{i>j}^{m} \frac{\sigma_i^2 + \sigma_j^2}{\sigma_i \sigma_j} u_i^2 u_j^2 - \sum_{i=1}^{m} u_i^4 - 2 \sum_{i>j}^{m} u_i^2 u_j^2 + R_2 \\ &= \sum_{i>j}^{m} \frac{\left(\sigma_i - \sigma_j\right)^2}{\sigma_i \sigma_j} u_i^2 u_j^2 + R_2 \\ &= R_1 + R_2, \end{aligned}$$

where

$$R_1 = \sum_{i>j}^{m} \frac{(\sigma_i - \sigma_j)^2}{\sigma_i \sigma_j} u_i^2 u_j^2,$$

$$R_2 = \left(\sum_{i=1}^{m} \sigma_i u_i^2\right) \left[\sum_{i=1}^{k} \frac{2\varepsilon - \varepsilon^2}{1+(1-\varepsilon)^2} \left(\frac{1}{\alpha_i}\right) v_i + \sum_{i=2k+1}^{m} \frac{1}{\lambda_i} u_i^2\right].$$

Thus, if we take δ so as $0 < \delta \le \varepsilon < 1$, the above inequality brings the estimation

$$\begin{aligned} &(2-\delta)\left(w^\top B w\right) \frac{1}{2} w^\top \left(B^{-1} + B^{-\top}\right) w - \left(w^\top w\right)^2 \\ &\quad \ge R_1 + \left(\sum_{i=1}^{m} \sigma_i u_i^2\right) \left[\sum_{i=1}^{k} \frac{2\varepsilon - \varepsilon^2 - \delta\left[1+(1-\varepsilon)^2\right]}{1+(1-\varepsilon)^2} \left(\frac{1}{\alpha_i}\right) v_i + \sum_{i=2k+1}^{m} \frac{1-\delta}{\lambda_i} u_i^2\right] \\ &\quad = R_1 + \left(\sum_{i=1}^{m} \sigma_i u_i^2\right) \left[\sum_{i=1}^{k} \frac{(2-\varepsilon)(\varepsilon-\delta) + \delta\varepsilon(1-\varepsilon)}{1+(1-\varepsilon)^2} \left(\frac{1}{\alpha_i}\right) v_i + \sum_{i=2k+1}^{m} \frac{1-\delta}{\lambda_i} u_i^2\right] > 0. \end{aligned}$$

Hence we obtain

$$(2-\delta)\left(w^\top B w\right) \frac{1}{2} w^\top \left(B^{-1} + B^{-\top}\right) w > \left(w^\top w\right)^2,$$

and the desired result (4) follows from the negative definiteness of B and $(B^{-1} + B^{-\top})$.

Conversely, suppose that for some i the inequality $|\beta_i/\alpha_i| \ge 1$ holds. Considering a vector u having nonzero entries only u_{2i-1} and u_{2i}, it suffices to evaluate the case of $m = 2$ with

$$B = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$$

and $|\beta/\alpha| \ge 1$. Then, we have

$$\begin{aligned} \left(w^\top w\right)^2 &= \left(w_1^2 + w_2^2\right)^2, \\ w^\top B w &= \alpha\left(w_1^2 + w_2^2\right), \\ w^\top \left(B^{-1} + B^{-\top}\right) w &= \frac{2\alpha}{\alpha^2 + \beta^2}\left(w_1^2 + w_2^2\right). \end{aligned}$$

Hence, these identities and the assumption that $|\beta/\alpha| \ge 1$ lead to

$$\frac{\left(w^\top w\right)^2}{\left(w^\top B w\right) \cdot (1/2) w^\top \left(B^{-1} + B^{-\top}\right) w} = 1 + \left(\frac{\beta}{\alpha}\right)^2 \ge 2. \qquad \blacksquare$$

3. A STABILITY RESULT

Applying Lemma 2, we arrive at our main result.

THEOREM 2. *The numerical solution by Lambert's method of the problem with a normal matrix satisfying Assumption A converges to 0 monotonically in L_2-norm if the eigenvalues of the matrix fulfill the condition $|\beta_i/\alpha_i| < 1$ for $1 \le i \le k$.*

PROOF. Let l and L be the same quantities as in Lemma 2. From (3), we obtain the estimation

$$l \le \frac{f_n^\top A f_n}{f_n^\top f_n} \le L,$$

which implies

$$-\frac{1}{l} \leq h_n^* \leq -\frac{1}{L}. \tag{9}$$

Applying Lambert's method (2) to the given problem yields

$$\boldsymbol{y}_{n+1} = (\boldsymbol{I} + h_n \boldsymbol{A})\, \boldsymbol{y}_n.$$

Therefore,

$$\begin{aligned} \boldsymbol{y}_{n+1}^\top \boldsymbol{y}_{n+1} &= \boldsymbol{y}_n^\top (\boldsymbol{I} + h_n \boldsymbol{A})^\top (\boldsymbol{I} + h_n \boldsymbol{A})\, \boldsymbol{y}_n \\ &= \boldsymbol{y}_n^\top \boldsymbol{y}_n + h_n \boldsymbol{y}_n^\top \left(\boldsymbol{A}^\top + \boldsymbol{A}\right) \boldsymbol{y}_n + h_n^2 \boldsymbol{y}_n^\top \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{y}_n. \end{aligned}$$

Hence, the equation

$$\frac{\|\boldsymbol{y}_{n+1}\|_2^2}{\|\boldsymbol{y}_n\|_2^2} = 1 - T_n \tag{10}$$

holds, where

$$T_n = -\frac{(1/2) h_n \boldsymbol{y}_n^\top \left(\boldsymbol{A}^\top + \boldsymbol{A}\right) \boldsymbol{y}_n}{\boldsymbol{y}_n^\top \boldsymbol{y}_n} \left(2 + \frac{h_n \boldsymbol{y}_n^\top \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{y}_n}{(1/2) \boldsymbol{y}_n^\top \left(\boldsymbol{A}^\top + \boldsymbol{A}\right) \boldsymbol{y}_n}\right).$$

Since $\boldsymbol{f}_n = \boldsymbol{A}\boldsymbol{y}_n$, we have

$$T_n = \frac{-(1/2) h_n \boldsymbol{y}_n^\top \left(\boldsymbol{A}^\top + \boldsymbol{A}\right) \boldsymbol{y}_n}{\boldsymbol{y}_n^\top \boldsymbol{y}_n} \left(2 + \frac{h_n \boldsymbol{f}_n^\top \boldsymbol{f}_n}{(1/2) \boldsymbol{f}_n^\top \left(\boldsymbol{A}^{-1} + \boldsymbol{A}^{-\top}\right) \boldsymbol{f}_n}\right). \tag{11}$$

The definition of h_n implies

$$h_n \leq h_n^* = -\frac{\boldsymbol{f}_n \boldsymbol{f}_n}{\boldsymbol{f}_n \boldsymbol{A} \boldsymbol{f}_n},$$

which, along with the negativeness of $\boldsymbol{f}_n^\top (\boldsymbol{A}^{-1} + \boldsymbol{A}^{-\top}) \boldsymbol{f}_n$ and (4), yields

$$\begin{aligned} \frac{h_n \boldsymbol{f}_n^\top \boldsymbol{f}_n}{(1/2) \boldsymbol{f}_n^\top \left(\boldsymbol{A}^{-1} + \boldsymbol{A}^{-\top}\right) \boldsymbol{f}_n} &\geq \frac{h_n^* \boldsymbol{f}_n^\top \boldsymbol{f}_n}{(1/2) \boldsymbol{f}_n^\top \left(\boldsymbol{A}^{-1} + \boldsymbol{A}^{-\top}\right) \boldsymbol{f}_n} \\ &= \frac{-\left(\boldsymbol{f}_n^\top \boldsymbol{f}_n\right)^2}{\left(\boldsymbol{f}_n^\top \boldsymbol{A} \boldsymbol{f}_n\right) \cdot (1/2) \boldsymbol{f}_n^\top \left(\boldsymbol{A}^{-1} + \boldsymbol{A}^{-\top}\right) \boldsymbol{f}_n} > -2 + \delta. \end{aligned}$$

From equation (11), we thus have the bound

$$T_n > -\frac{(1/2) h_n \boldsymbol{y}_n^\top \left(\boldsymbol{A}^\top + \boldsymbol{A}\right) \boldsymbol{y}_n}{\boldsymbol{y}_n^\top \boldsymbol{y}_n} \delta,$$

and, together with (3), we furthermore obtain

$$T_n > -\delta h_n L.$$

CASE 1. $h_0 < h_n^*$. Then $h_n = h_0$ and $T_n > -\delta h_0 L =: \gamma_1 > 0$. (Note that $\gamma_1 < -\delta h_n^* L \leq \delta < 1$ by the inequality (9).)

CASE 2. $h_0 \geq h_n^*$. Then $h_n = h_n^*$ and $T_n > -\delta h_n^* L \geq \delta L/l =: \gamma_2 > 0$ by the inequality (9). (Again note that $\gamma_2 < 1$.)

If we take $\gamma = \min(\gamma_1, \gamma_2)$, then $T_n > \gamma > 0$ holds for all n, and the result is established. ∎

This theorem indicates that Lambert's method is acceptable for a linear system with real normal matrix coefficients if all of its eigenvalues lie in the sector of the angle within ±45° from negative real axis. We note that, in the case where the initial steplength, h_0, is smaller than the amount which is determined by the stability condition for Euler's rule, Lambert's method (2) is always stable regardless of the size of $|\beta_i/\alpha_i|$ because h_n is always chosen to be smaller than that amount.

4. NUMERICAL EXPERIMENTS

We will confirm Theorem 2 through a numerical example. We consider the following constant coefficient linear problem whose Jacobian matrix is real and normal.

$$\begin{bmatrix} y_1'(x) \\ y_2'(x) \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} \qquad \text{for } 0 \le x \le 10,$$
$$\begin{bmatrix} y_1(0) \\ y_2(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

where α and β are real constants. The eigenvalues of this problem are $\alpha \pm i\beta$, and the theoretical solution is

$$\begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} = \begin{bmatrix} [\cos(\beta x) + \sin(\beta x)] \exp(\alpha x) \\ [\cos(\beta x) - \sin(\beta x)] \exp(\alpha x) \end{bmatrix}.$$

We applied the method (2) to this IVP with $\alpha = -10$ and $\beta = 8, 10, 12$. Numerical calculations are summarized in Table 1, where "Euler's Condition" means the maximum allowable steplength when Euler's rule is applied to this IVP. Error in Table 1 is the maximum of absolute error at the endpoint, $t = 10$. (The numerical solutions at $t = 10$ were computed by a linear interpolant over the last step.)

Table 1. Numerical results.

Eigenvalues	Euler's Condition	h_0	Error
$-10 \pm 8i$	0.12	0.1	2.04×10^{-10}
		0.2	2.04×10^{-10}
$-10 \pm 10i$	0.10	0.08	4.81×10^{-11}
		0.12	1.00
$-10 \pm 12i$	0.082	0.07	1.00×10^{-7}
		0.09	1.05×10^{4}

From Table 1, we can observe that Lambert's method gives stable solutions for every initial steplength h_0 if the conditions of Theorem 2 are satisfied. But the numerical solution loses its stability if the conditions of Theorem 2 are violated and the initial steplength is larger than the "Euler's Condition." This fact endorses our Theorem 2. We note that when $\alpha = \beta$ and h_0 is larger than the "Euler's Condition," equation (10) indicates that the magnitude of the numerical solution does not change since, in this case, $T_n = 0$ (if $h_n^* \le h_0$). This is verified by the numerical result with $\beta = 10$ and $h_0 = 0.12$.

REFERENCES

1. J.D. Lambert, A stable sequence of steplengths for Euler's rule applied to stiff systems of differential equations, *Computers Math. Applic.* **12B** (5/6), 1141–1151 (1986).
2. R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, London, (1990).